ARI Research Note 89-35

# Improvement in Command Group Performance as a Result of Diagnostic Feedback and CATTS Training

Gary S. Thomas, Herbert F. Barber,
and Ira T. Kaplan
U.S. Army Research Institute



Field Unit at Fort Leavenworth, Kansas
Stanley M. Halpin, Chief

Systems Research Laboratory
Robin L. Keesee, Director

June 1989

United States Army
Research Institute for the Behavioral and Social Sciences

# U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

## A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

**EDGAR M. JOHNSON**
**Technical Director**

**JON W. BLADES**
**COL, IN**
**Commanding**

Technical review by

Lloyd M. Crumley



### NOTICES

**DISTRIBUTION**: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

**FINAL DISPOSITION**: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

**NOTE**: The views, opinions, and findings in this report are those of the author(s) and should not to be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

# REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

| | |
|---|---|
| 1a. REPORT SECURITY CLASSIFICATION<br>Unclassified | 1b. RESTRICTIVE MARKINGS<br>-- |
| 2a. SECURITY CLASSIFICATION AUTHORITY<br>-- | 3. DISTRIBUTION / AVAILABILITY OF REPORT<br>Approved for public release; distribution is unlimited. |
| 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE<br>-- | |
| 4. PERFORMING ORGANIZATION REPORT NUMBER(S)<br>ARI Research Note 89-35 | 5. MONITORING ORGANIZATION REPORT NUMBER(S)<br>-- |
| 6a. NAME OF PERFORMING ORGANIZATION<br>U.S. Army Research Institute Field Unit at Fort Leavenworth<br>6b. OFFICE SYMBOL (If applicable)<br>PERI-SL | 7a. NAME OF MONITORING ORGANIZATION<br>-- |
| 6c. ADDRESS (City, State, and ZIP Code)<br>P.O. Box 3407<br>Fort Leavenworth, KS 66027-0347 | 7b. ADDRESS (City, State, and ZIP Code)<br>-- |
| 8a. NAME OF FUNDING / SPONSORING ORGANIZATION<br>U.S. Army Research Institute for the Behavioral and Social Sciences<br>8b. OFFICE SYMBOL (If applicable)<br>PERI-S | 9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER<br>-- |
| 8c. ADDRESS (City, State, and ZIP Code)<br>5001 Eisenhower Avenue<br>Alexandria, VA 22333-5600 | 10. SOURCE OF FUNDING NUMBERS<br>PROGRAM ELEMENT NO. 62785A; PROJECT NO. 791; TASK NO. 1301; WORK UNIT ACCESSION NO. H2 |

11. TITLE (Include Security Classification)
Improvement in Command Group Performance as a Result of Diagnostic Feedback and CATTS Training

12. PERSONAL AUTHOR(S)
Thomas, Gary S.; Barber, Herbert F.; and Kaplan, Ira T. (ARI)

| 13a. TYPE OF REPORT | 13b. TIME COVERED | 14. DATE OF REPORT (Year, Month, Day) | 15. PAGE COUNT |
|---|---|---|---|
| Final | FROM 1982 TO 1983 | 1989, June | 31 |

16. SUPPLEMENTARY NOTATION This report was originally published in 1983 as an internal ARI document not available for general distribution. It has been made available because it contains data now of general interest to C2 researchers.

| 17. COSATI CODES | | | 18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) | |
|---|---|---|---|---|
| FIELD | GROUP | SUB-GROUP | Command | Battalion |
| | | | Control | Performance |
| | | | Command post | Evaluation |

19. ABSTRACT (Continue on reverse if necessary and identify by block number) Five battalion command groups participated in four 1-day (CATTS) exercises. Performance on the first and last exercises was compared. Researchers measured information reception and transmission by group members during planning (information flow questionnaire); information exchange by staff members during battle execution (probes); degree of success on the simulated battlefield (mission accomplishment scores); and BCG ARTEP performance as assessed by CATTS controllers, player-controllers, and the players themselves (subjective ratings). Feedback concerning performance was provided after each day's training. All measures of performance increased significantly from first day to fourth day. Battalion command group performance on an information flow measure during both planning and execution and performance on the battlefield improved as a result of CATTS training. Since performance did not increase in an earlier similar experiment where no feedback was provided, it is likely that the feedback sessions contributed substantially to the increase in performance observed in this research. The subjective ratings of ARTEP performance observed in the current research were greater than those in the previous research.

| | |
|---|---|
| 20. DISTRIBUTION / AVAILABILITY OF ABSTRACT<br>☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS | 21. ABSTRACT SECURITY CLASSIFICATION<br>Unclassified |
| 22a. NAME OF RESPONSIBLE INDIVIDUAL<br>Lloyd M. Crumley | 22b. TELEPHONE (Include Area Code)<br>(913) 684-4933<br>22c. OFFICE SYMBOL<br>PERI-SL |

DD Form 1473, JUN 86 *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

## FOREWORD

The Fort Leavenworth Field Unit of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) conducts a systems and training research program in support of the Combined Arms Center. This report describes the second of two experiments conducted by the Field Unit contributing to the Army Training Battle Simulation System (ARTBASS) Training Development Study. The report represents an interim analysis of the data to respond in a timely fashion to requirements of the research sponsors and ARTBASS proponents, the Battle Simulation Directorate, at Fort Leavenworth.

IMPROVEMENT IN COMMAND GROUP PERFORMANCE AS A RESULT OF DIAGNOSTIC FEEDBACK AND CATTS TRAINING

## EXECUTIVE SUMMARY

### Requirement:

The Combined Arms Center at Fort Leavenworth is responsible for developing command and control ($C^2$) training systems. One of the most promising recent approaches is the application of automation in computer-driven $C^2$ training systems. The most advanced of these systems is the Combined Arms Tactical Training Simulator (CATTS), which has been used as a training vehicle and training development test bed. Based on information gathered during the use and refinement of CATTS, a follow-on system, the Army Training Battle Simulation System (ARTBASS), is now being developed. To demonstrate that ARTBASS will be effective for training battalion command groups (BCGs) in $C^2$ behaviors, the ARTBASS Test Integration Working Group decided that a Training Development Study (TDS) could be conducted using CATTS in lieu of ARTBASS due to the similarity between the systems.

In support of CATTS/ARTBASS development and the TDS, the efforts of the ARI Field Unit at Fort Leavenworth have included the refinement of $C^2$ measurement techniques and the identification of CATTS system and scenario characteristics that impact on command group $C^2$ performance so that changes in $C^2$ performance as a function of CATTS training can be documented. Improved diagnostic and feedback procedures were implemented in the research to enhance the training effectiveness of CATTS. The present experiment assessed the effects of CATTS training, supplemented by an ARI-developed performance diagnosis and feedback package, on the $C^2$ performance of BCGs.

### Procedure:

Five battalion command groups (players), three mechanized infantry, and two armor battalions participated in four 1-day CATTS exercises between December 1982 and March 1983. The first and last exercises were designated as pretest and posttest exercises, respectively, and included force/delay missions on different portions of Fulda Gap terrain. The intervening training exercises were delay and movement-to-contact conducted on Ft. Irwin or Sinai terrain. Researchers measured information reception and transmission by group members during planning (information flow questionnaire); information exchange by staff members during battle execution (probes); degree of success on the simulated battlefield (mission accomplishment scores); and BCG ARTEP performance as assessed by CATTS controllers, player-controllers, and the players themselves (subjective ratings).

Measures of performance were collected on the pretest exercises, and with the exception of mission accomplishment scores, the results were presented to command group members in individual feedback sessions conducted by CATTS controllers and ARI personnel before the training exercises. Those areas needing improvement were addressed by the players during training exercises. At the conclusion of pretest and training exercises, significant events from each day's battle were presented to the players via an "instant replay" capability of the CATTS computer. A CATTS controller conducted the battle replay sessions so that strengths and weaknesses of battle procedures could be identified. Players and player-controllers also conducted after-action reviews. Performance measures were again collected on the posttest exercises.

Findings:

All measures of performance increased significantly from pre- to posttest exercises. BCG performance on information flow during both planning and execution and performance on the battlefield improved as a result of CATTS training. Since BCG performance on information flow during planning and on the simulated battlefield did not increase as a result of CATTS training when detailed feedback was not provided in previous research (Thomas, Barber, & Kaplan, 1983), it is likely that the feedback sessions contributed substantially to the increase in performance observed in this research. In addition, the increase in subjective ratings of ARTEP performance observed in the current research were greater than those in the previous research.

Utilization of Findings:

The degree to which these findings apply to ARTBASS is limited by the degree to which ARTBASS resembles CATTS as it was used in the current research. The research implies that, without a mechanism for detailed command group performance diagnosis and feedback, ARTBASS may be less than optimally effective for training BCGs in the field.

IMPROVEMENT IN COMMAND GROUP PERFORMANCE AS A RESULT OF DIAGNOSTIC FEEDBACK AND CATTS TRAINING

CONTENTS



LIST OF TABLES

Page

LIST OF TABLES (Continued)



LIST OF FIGURES

# IMPROVEMENT IN COMMAND GROUP PERFORMANCE AS A RESULT OF DIAGNOSTIC FEEDBACK AND CATTS TRAINING

## INTRODUCTION

The advent of automation has generated new and expanded possibilities for the development of training vehicles to meet the Army's needs in the 1980's and 90's. One of the most promising uses of automation is in the area of command and control ($C^2$) training. Specifically, the Combined Arms Tactical Training Simulator (CATTS), a computer driven, free-play $C^2$, battle simulation system was developed as a training vehicle and training development test bed. Based on information gathered during the development and refinement of CATTS, a follow-on system, the Army Training Battle Simulation System (ARTBASS) is now being developed.

Prior to fielding, it is necessary to demonstrate that ARTBASS is, in fact, effective for training command groups in $C^2$ behaviors. The development schedule for ARTBASS severely limits the amount of time available to determine its training effectiveness, but since ARTBASS in essence grew from CATTS, there is a close similarity between the two. Therefore, the ARTBASS Test Integration Working Group decided that a Training Development Study (TDS) should be conducted using CATTS in lieu of ARTBASS. The generalizability of TDS results to ARTBASS is, of course, limited by the degree to which the two systems are actually similar.

Previous research (Thomas, Barber, and Kaplan, 1983) identified several system and scenario characteristics (e.g., combat ratio, mission type, weather, etc.) which significantly influenced the success achieved by battalion command groups (BCGs) on the simulated battlefield, as measured by battle outcomes. In that research, performance feedback to BCGs was limited to an after-action review conducted by the players themselves at the conclusion of each exercise day. Neither measures of battlefield performance, nor measures of inter- and intra-staff communication during planning increased with CATTS training. Subjective ratings of BCG performance on a set of ARTEP subtasks provided by company commanders did not increase as a function of CATTS training. On the other hand, ratings by CATTS controllers and the players themselves increased from the first day to the third and fourth days' exercises.

In support of ARTBASS development and the TDS, the efforts of the ARI Field Unit at Fort Leavenworth included the refinement of $C^2$ measurement techniques and the development of a diagnostic and feedback module, so that changes in $C^2$ performance as a function of exposure to CATTS could be documented. Therefore, the present research assessed the effect of CATTS training, supplemented by performance diagnosis and feedback, on a variety of BCG $C^2$ measures.

## METHOD

### Participants

Five battalion command groups each participated in four one-day CATTS training exercises between December 1982 and March 1983. Three of the command groups (players) were armor battalions, and two were mechanized infantry. Table 1 characterizes the typical composition of each command group, where each key position (e.g., battalion commander, S1, S2, S3, and S4) was occupied by the individual who normally filled that positions, i.e., incumbant. Several supporting members (not standardized) also participated in exercises.

### Experimental Design

A pretest, training, post-test design was used to assess the combined effects of performance feedback and CATTS training on several measures of performance. Performance measures were collected on pre and post-test exercises, where BCGs conducted covering force/delay missions on two similar and adjacent parts of Fulda Gap terrain. The portion of terrain used were partially counter-balanced, so that three BCG's operated on the northern portion on the pretest and the southern portion on the post-test, and the order of terrain presentation was reversed for the remaining two BCGs. Other potentially important variables such as initial combat ratio, weather, and extent of jamming were equivalent for each exercise. Performance feedback on the measures collected during the pretest was presented to the BCGs on the morning after the pretest exercise. Prior to the post-test exercise, each BCG conducted two training exercises where delay and movement to contact missions were executed on desert terrains. At the conclusion of each exercise day there was an after-action review and a replay of the battle conducted by a CATTS controller. The schedule of events for the exercises appears in Table 2.

### Training System

Simulation. The battlefield environment was simulated by the CATTS computer, which provided a computer-driven exercise to train maneuver-battalion commanders and their staffs in the control and coordination of combined-arms operations. CATTS simulated the actions of units in combat, moved elements on and about the battlefield, calculated intervisibility and detection between forces, calculated weapon-to-target ranges, and the effects of weapons employment. It also maintained the status of personnel, equipment, ammunition, and fuel for friendly and enemy forces. Speed of maneuver, line of sight, and weapons effects were affected by changes in weather, terrain contour, soil type, suppressive fires, and personnel and equipment status. Given line of sight, engagements among maneuver weapon systems were automatic.

Table 1

Battalion Command Group Performance

Battalion Commander

S1

S4

S1 or S4 NCO

S1 or S4 RTO

S2

S2 NCO

S3

S3 Air

S3 NCO

S3 RTO

Company Commanders (2 tank, 2 line)

Fire Support Officer

Fire Support NCO

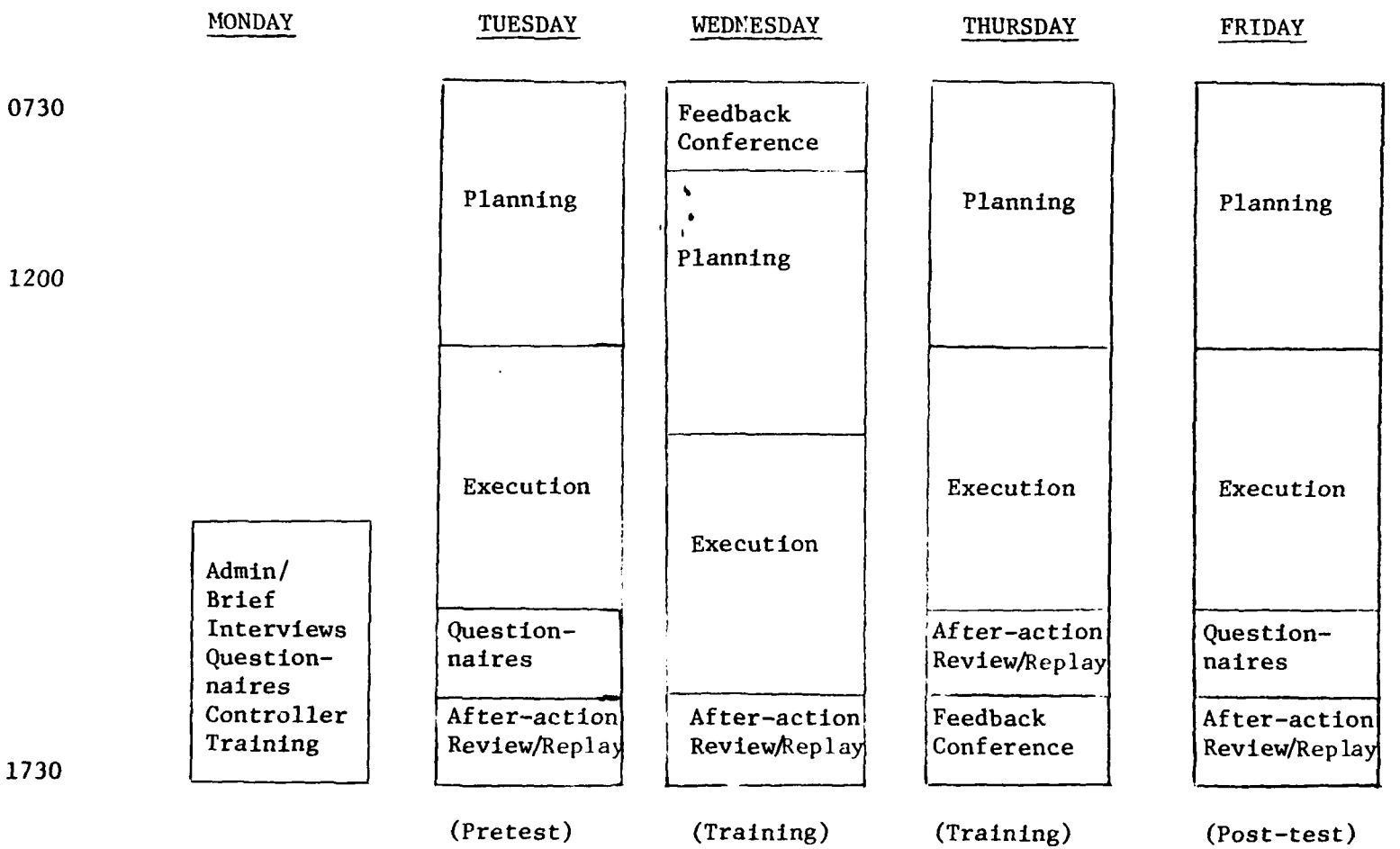
Fist Chief (two)

Air Liaison Officer (Air Force)

Forward Air Controller (Air Force)

Table 2

General Schedule Of Events



The CATTS exercises were conducted in a real-time, free-play mode. Within the prescribed tactical situation, the battalion commander could employ battalion assets in any manner deemed appropriate. The only constraints were the assets available to the battalion and the actions of the enemy commander. Deployment of enemy assets were consistent across exercises, but, in accordance with threat doctrine, the threat controller made minor tactical adjustments to counter unique situations created by friendly force operations.

In this research, the command group, except the S1 and S4, occupied a simulated tactical operations center (TOC); the S1 and S4 were in another area, designated as the combat trains. The players (the battalion command group) in both areas were provided with communications equipment normally found in a maneuver battalion. They could communicate with higher, lower, and adjacent units (played by controllers) in any manner consistent with Army procedure and with the simulated location of the various units: face-to-face, by telephone, by radio, and by written message.

Figure 1 illustrates the communication among the players, the controllers, and the computer. Most communication took place by radio and telephone. The BCGs had seven radio nets (actually hard-wired) with appropriate alternate frequencies. The nets included the following: the brigade command, brigade intelligence, brigade administration/logistics, battalion command, battalion administration/logistics, and air support nets. In addition, the command group also had a RATT (radioteletype) unit and field telephones, when appropriate. The sounds of enemy jamming, battle, and engine and generator noise were generated during the exercise to enhance the realism of the experience.

Controllers. A team of controllers, permanently assigned to CATTS, mediated between the players and the computer. The control team consisted of a chief controller, who also played the role of brigade commander, and brigade S1, S2, and S3, and S4 controllers. In addition, a fire support controller, a direct air support controller, and a threat controller were also present. Three additional controllers identified as interactors, input orders into the computer at three control consoles: (a) the command and control interactor input orders from the battalion command group via company commanders to the maneuver units modeled in the computer, (b) the fire support interactor input orders to the friendly artillery and air support units, and (c) the threat interactor input actions directed by the threat controller.

Player-Controllers. Each command group brought along its company commanders and fire support representatives to serve as player-controllers. They received orders from battalion and translated them into subordinate unit maneuvers for input into the computer by the interactors. In addition, they also received battle status reports from the computer and relayed that information back to the command group in the form of situation reports and spot reports.

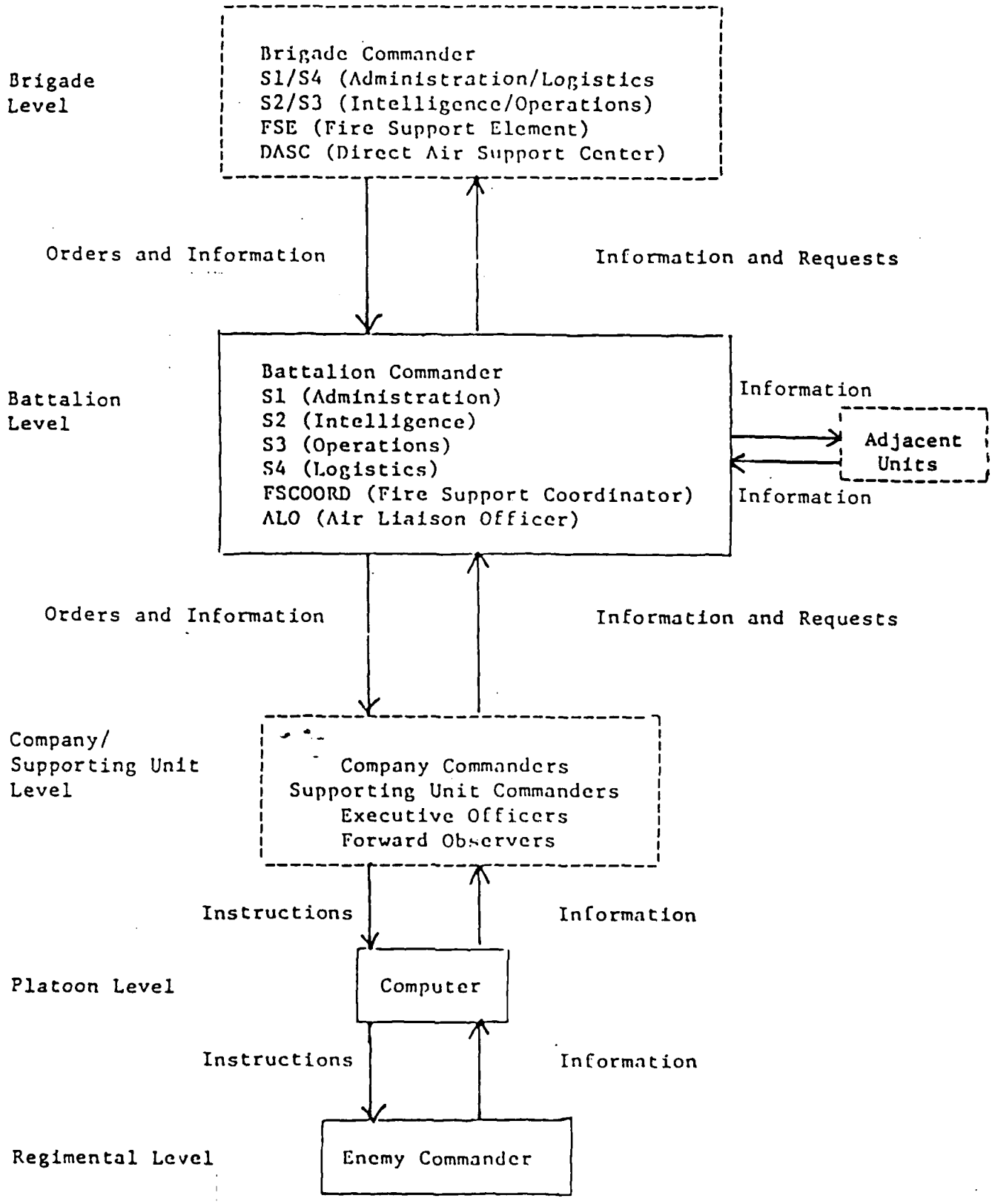



Figure 1. Communication between controller and player positions in CATTS. Controller positions are inclosed by broken lines.

## Feedback System

In the battle replay sessions at the end of each exercise day, CATTS controllers replayed significant events on a video display of the terrain, which showed the position and strength of friendly and enemy forces during the battle, and the players and player-controllers presented their perceptions of the battle in after-action reviews. Feedback conferences were held on the morning following the pretest exercise. ARI personnel provided the battalion commander with results from an information flow questionnaire, described below, and with ratings of BCG performance on a short list of ARTEP tasks. Problem areas were identified and solutions were recommended. The remaining staff members (S1, S2, S3, S4, FSO, and ALO) met with their brigade controller counterparts and discussed problems in command and control processes identified in the pretest exercise. Discussions included, but were not confined to, performance on ARTEP tasks and to the group's ability to transmit important information during mission execution (probes). Problem areas were identified and tentative solutions recommended by controllers. At the conclusion of the third exercise day, feedback conferences were again conducted with the same individuals. Discussions focused on problems that were resolved, new problem areas, and recommended solutions to problems.

## Performance Measures

The following measures were intended to assess the command group's ability to transmit important information during planning (information flow questionnaire) and execution (probes), to perform ARTEP tasks (subjective ratings), and to perform on the simulated battlefield (mission accomplishment scores).

Information Flow. At the beginning of each pretest and post-test exercise, the principal members of the battalion staff, the S1, S2, S3, S4, FSO, and ALO were briefed separately by their brigade counterparts. In these briefings, certain unique items of information were presented to each member. Then the battalion commander and his staff worked together for 3 to 4 hours to develop a plan which they presented to their company commanders during the battalion operations order briefing. Subsequently, the command group members and company commanders answered multiple-choice questions based upon the unique information originally presented in the brigade briefing. Examples of questions are presented in Appendix A. There were two parallel forms of the questionnaire, one for the pretest and one for the posttest. The order of presentation of the questionnaires was counterbalanced across the five units. The responses were analyzed to provide three measures of information flow:

1. Reception of required information that was presented to the staff member during the brigade briefing. The percent of these items answered correctly was the individuals direct reception score.

2. Reception of required information that the command group members should have received indirectly from those members who received it directly from brigade. The percent of such items answered correctly was the intra-group communication score.

3. Reception of information required by the company commanders that should have been transmitted to them by members of the battalion command group. The percent of such items answered correctly measured communication to the company commanders during the battalion operations order briefing.

Probes. In a procedure similar to the information flow technique, a series of preplanned "probes" were inserted into the exercise during the execution phase. The probes were designed to assess coordination, communication, and information processing behaviors within the command group. Probes were events that could realistically occur during an operation. Some probes were used in both the pretest and the post-test, because they occurred as a natural outgrowth of the exercise, e.g., asking for a situation report or an estimate of enemy intentions; others were used only once in either the pretest or the post-test, because they might seem artificial if they were repeated, e.g., indications that an adjacent unit was suffering a chemical attack. The unique probes were counterbalanced across command groups to control for difficulty. The number of probes was limited, so as not to interfere with the controllers' primary functions or the flow of the exercise. Each of five controllers (S1, S2, S3, S4, and FSO) inserted four to six probes in each pre- or post-test exercise and evaluated the timeliness, completeness, and accuracy of the response. An example of a probe appears in Appendix B.

Mission Accomplishment Scores. Procedures for deriving mission accomplishment scores that are based on BCG achievement of the objectives of a covering force/delay mission are described in Thomas and Cocklin (1983). That research described how performance on separate mission objectives could be combined into an overall assessment of battlefield mission accomplishment.

The current research used the same set of mission objectives of a covering force/delay mission as did the previous work. These measures included (1) relative losses between opposing forces (surviving maneuver force ratio differential), (2) whether or not friendly forces were combat effective (50% of initial strength) at the conclusion of battle, (3) the quality of friendly intelligence gathering (estimates of enemy strength, location, rate of advance, and likely course of action), (4) the depth of OPFOR advance during the attack, and (5) friendly force location with respect to the OPFOR and the MBA.

The values for each of these measures were collected on all pre- and post-test exercises were presented to four expert military judges. The judges were retired officers, whose ranks ranged from lieutenant colonel to brigadier general. Each had extensive experience in combat or combat modeling. Based on the measures of mission objectives obtained for each

exercise, the judges gave an overall mission accomplishment score to each set of battle outcomes. Scores could range from 100, indicating perfect mission accomplishment, to 0, indicating total failure. A value of 50 was the mid-point on the mission accomplishment scale.

ARTEP Ratings. During pre- and post-test CATTS exercises, ratings of BCG performance were obtained from CATTS controllers, from the players themselves, and from player-controllers (company commanders, and FIST). Items were based on a modified list of tasks taken from the command group/ staff module of ARTEP 71-2, (Army Training and Evaluation Program . . . 71-2, 1977). These tasks (Appendix C) were selected on the basis of previous ARI research (Barber and Kaplan, 1979; Kaplan and Barber, 1979; Barber and Solick, 1980; and Thomas, Barber, and Kaplan, 1983). Ratings were obtained from a one to nine scale, where one corresponded to poor and nine to outstanding performance. In addition to rating group performance, the controllers also rated the performance of the principal group members.

## RESULTS AND DISCUSSION

In general, battalion command group performance increased significantly from pretest to post-test exercises on all measures as a result of diagnostic feedback and CATTS training. The following provides a detailed description of the results for each performance measure.

### Information Flow

The information-flow questionnaire measured three stages in the process of communication: (1) communication from brigade to battalion, (2) communication within a battalion command group and (3) communication from battalion to company. Prior to analysis of the information flow scores, the split-half reliability of the information-flow questionnaire based on a sample of 46 questionnaires was calculated and found to be a respectable .82. Further, the controllers' identification of required information was validated by the receivers' (players and player-controllers) ratings of the importance of the items. The mean importance rating of all required items was 7.1, approximately midway between 5 (moderately important) and 9 (essential).

Table 3 presents the percentages of information received in each stage of communication. One-tailed t-tests were applied to the matched pretest/ post-test scores for each of the five command groups. In all three stages the communication scores increased significantly from pretest to post-test. The greatest improvement (25%) occurred in the reception of information by the company commanders.

The controllers were also asked to observe the operations order briefings provided to the company commanders. Each controller was asked to record the number of required items briefed, to rate how good the briefings were, and to comment on the briefings. An average of 13.2 items were briefed in the pretest, versus 17.6 items in the post-test, an increase that was not

Table 3

Effect of Training with Feedback
on the Mean Percent of Required Information Received

| Stage of Communication | Pretest | Post-test | t |
|---|---|---|---|
| Brigade to Battalion | 85 | 89 | 2.21* |
| Within the Bn Cmd Grp | 51 | 66 | 4.97** |
| Battalion to Co Cdrs | 42 | 67 | 4.96** |

* $p < .05$ **$p < .01$ df = 4 one-tailed test for matched pairs

Table 4

Mean Percent of Probes Correct on
Pretest and Post-test by Staff Sections

| | Pre | Post |
|---|---|---|
| S1 | 87 | 90 |
| S2 | 72 | 86 |
| S3 | 77 | 87 |
| S4 | 51 | 79 |
| FSO | 30 | 60 |
| $\overline{X}$ | 63.4 | 80.4 |

statistically significant. Further, the average rating of the briefings increased from 4.5 (less than good) in the pretest, to 5.9 (half way between good and very good) on the post-test. This improvement was statistically significant at the .01 level ($t = 3.96$). Finally, the controllers commented on briefings that were particularly good or particularly bad. The most frequent comments (46%) referred to the completeness of the presentation. Typical unfavorable comments were that presentation was incomplete, sketchy, or omitted important information. Favorable comments were that it was thorough and covered most pertinent information. The next largest category (28%) concerned delivery. Critical comments were that the presentation was rushed, hurried, unsure, or unclear. Favorable comments were concise, clear, or good delivery. Some comments (8%) mentioned organization, saying the presentation was either poorly organized or well organized. In every category the percentage of favorable comments increased from the pretest to the post-test. Seventy-six percent of all the comments on the post-test were favorable, compared to 27% in the pretest.

## Probes

Each probe was designed to elicit information-processing by at least one staff member, so such behavior could be sampled during mission execution. Some probes were more effective than others in eliciting these behaviors. Probes that were not responded to more than 40% of the time were considered ineffective and were excluded from further analysis. The remaining probes were scored correct or incorrect depending upon the completeness and accuracy of the responses, as assessed by the controllers who administered the probes. Percent of probes correct was calculated for each staff section (S1, S2, S3, S4, FSO) on each exercise day. These percentages were cast into a 2 x 5 repeated measures ANOVA, to determine whether the command groups improved from pretest to post-test. It was found that the command groups' responses improved ($F_{1,4} = 4.68$, $p < .05$) as indicated in Table 4. There was also a significant difference in percent correct between staff sections ($F_{1,16} = 4.03$, $p < .02$), where all staff sections performed better than the FSOs (HSD = 4.23, $p < .05$). This latter finding could mean either that FSO performance was inferior to that of other staff sections or that the FSO probes were more difficult than the other probes. Finally, in a separate analysis, no significant differences among command groups were observed in responses to probes ($F_{1,4} = .02$).

## Mission Accomplishment

The degree to which judges agreed on their ratings of battlefield mission accomplishment was determined by correlating those scores among the four judges as indicated in Table 5. There was very high agreement among judges 1, 2, and 3, and moderately high agreement between these judges and judge 4. Since there was generally good agreement among judges' ratings, all mission accomplishment scores (Table 6) were included in an analysis to determine if these scores improved as a function of CATTS training.

Table 5

Interrater Agreement on Mission Accomplishment Scores

| | $J_1$ | $J_2$ | $J_3$ | $J_4$ |
|---|---|---|---|---|
| $J_1$ | --- | .951 | .995 | .757 |
| $J_2$ | | --- | .942 | .748 |
| $J_3$ | | | --- | .751 |
| $J_4$ | | | | --- |

Table 6

Mean Mission Accomplishment Scores

*for Units and Days*

| | PRE | POST |
|---|---|---|
| Unit 1 | 66.50 | 24.50 |
| Unit 2 | 23.75 | 80.50 |
| Unit 3 | 24.25 | 70.75 |
| Unit 4 | 52.75 | 80.00 |
| Unit 5 | 41.25 | 79.75 |
| $\overline{X}$ | 41.70 | 67.10 |

A 2 X 5 repeated measures ANOVA indicated that battalions tended to improve on battlefield mission accomplishment from pre- to post-test ($F1,2 = 15.88$, $P < .028$). It was also determined that some battalions performed better on the simulated battlefield regardless of exercise day as indicated by a significant unit effect ($F_{4,12} = 3.84$, $P < .031$). Finally, as indicated in Table 6, one unit (unit 1) performed more poorly on the post-test than on the pretest. This is reflected in a significant unit by exercise day interaction ($F_{4,12} = 16.68$, $P < .001$).

ARTEP Ratings

Prior to testing for changes in ARTEP ratings from pretest to post-test, the ratings were analyzed to determine if the raters were (a) discriminating among items and (b) in agreement in their ratings of ARTEP performance.

Item Discrimination. Ratings of ARTEP performance made by controllers, players, and player-controllers were correlated between items for each rater and then combined for each rater group. These interitem correlations were highly significant for ($p < .002$) for all comparisons. The median interitem correlations for controllers, players, and player-controllers were .80, .71, and .65, respectively, indicating that raters typically did not discriminate among the various aspects of player performance on ARTEP tasks. That is to say, if command group performance was rated high on one ARTEP task, performance on all other ARTEP tasks tended to be rated high.

Ratings made by each of the three sets of raters were also subjected to factor analysis to determine what underlying variables (factors) could account for the variance in ratings. In each analysis, a one-factor solution accounted for 100% of the variance in ratings, providing further evidence that raters were responding to all items in a similar fashion and not discriminating among the ARTEP tasks. Therefore, in subsequent data analyses an average ARTEP rating was used for each rater on each exercise day. This was achieved by calculating an arithmetic mean across all ARTEP items for each rater on each day.

Interrater Agreement. The degree to which observers agreed in their ratings of command group, ARTEP performance could only be determined for controllers, since the too little data was available from players, and player-controllers. Interrater agreement was determined by correlating the average ARTEP ratings made by controllers on each exercise day. The highest agreement was between the chief controller and the S3 controller ($r = .90$). The S1 and S4 were also in high agreement in their ratings ($r = .88$). The lowest degree of relationship between ratings were between the FSO and the S1 controller ($r = .25$), and between the FSO and the S4 ($r = .29$). These results may reflect different perspectives on the exercises. For example, the chief controller and the S3 observed from the brigade control station in the control room, the S1 and S4 sat next to each other at another station, and the FSO observed from a third station. Overall, the median correlation among raters was .63, indicating a moderately high level of agreement among controllers in their ratings of command group performance.

Command Group Performance Change. Separate analyses were performed on ARTEP ratings for each group of raters to determine if post-test performance was perceived as being higher than pretest performance. Since controllers observed all exercises, more detailed analyses of their ratings was possible. All ratings of BCG performance on the set of ARTEP tasks were averaged across items for each rater on each exercise day.

Controller ratings of BCG performance were analyzed in a 2 X 5 repeated measures ANOVA to determine any pre- to post-test differences in ratings of command group performance. As indicated in Table 7 and verified by the analysis, command group performance was rated higher on the post-test exercises ($F_{1,3} = 30.46$, $p < .0001$), but no command groups were rated significantly higher than any others ($F_{4,12} = .99$). Also there was no significant exercise day-by-unit interaction ($F_{4,12} = 1.42$). That is to say, that controllers did not perceive any difference in the magnitude of the increase in ARTEP performance from pre-to post-test for any unit compared to the others.

Since differences in player and player-controllers ratings between units and unit-by-day interactions were not of interest, only potential differences in ratings from pre- to post-test were analyzed using dependent t-tests. As indicated in Figure 2, players perceived an increase in their ARTEP performance from the pretest to the post-test exercises ($t = 2.44$, $df = 4$, $p < .05$). Player-controllers also indicated that BCG ARTEP performance increased as a result of CATTS training as indicated in Figure 2 ($t = 6.61$, $df = 4$, $p < .005$). In conclusion, all raters thought command group performance increased significantly as a result of CATTS training and diagnostic feedback.

Ratings of Individual Staff Members. In addition to rating command group performance, controllers also rated the overall performance of each individual staff member whom they observed during CATTS exercises. All ratings were transformed to Z-scores for each observer, to control for rater response bias. These data were then analyzed to determine if there were differences in performance between command groups, or differences in performance on the pre- versus the post-test. No significant differences in performance were observed between staff sections ($F_{5,20} = 1.19$); for example, S3 players performed no better or worse than other group members. In addition, there was no significant difference in performance among members of different command groups ($F_{4,20} = 1.30$). Finally, controller ratings of individual staff members did increase signficantly from the pretest exercise to the post-test. ($F_{1,20} = 131.82$, $p < .001$). This increase in the ratings of individual performance is consistent with the pre- to post-test increase in command group ARTEP ratings described above.

## CONCLUSIONS

Command group performance, as assessed by all measures developed in previous research and refined for the current experiment, improved significantly from the first day to the fourth day of CATTS training. All staff members improved in their ability to transmit and receive information during the

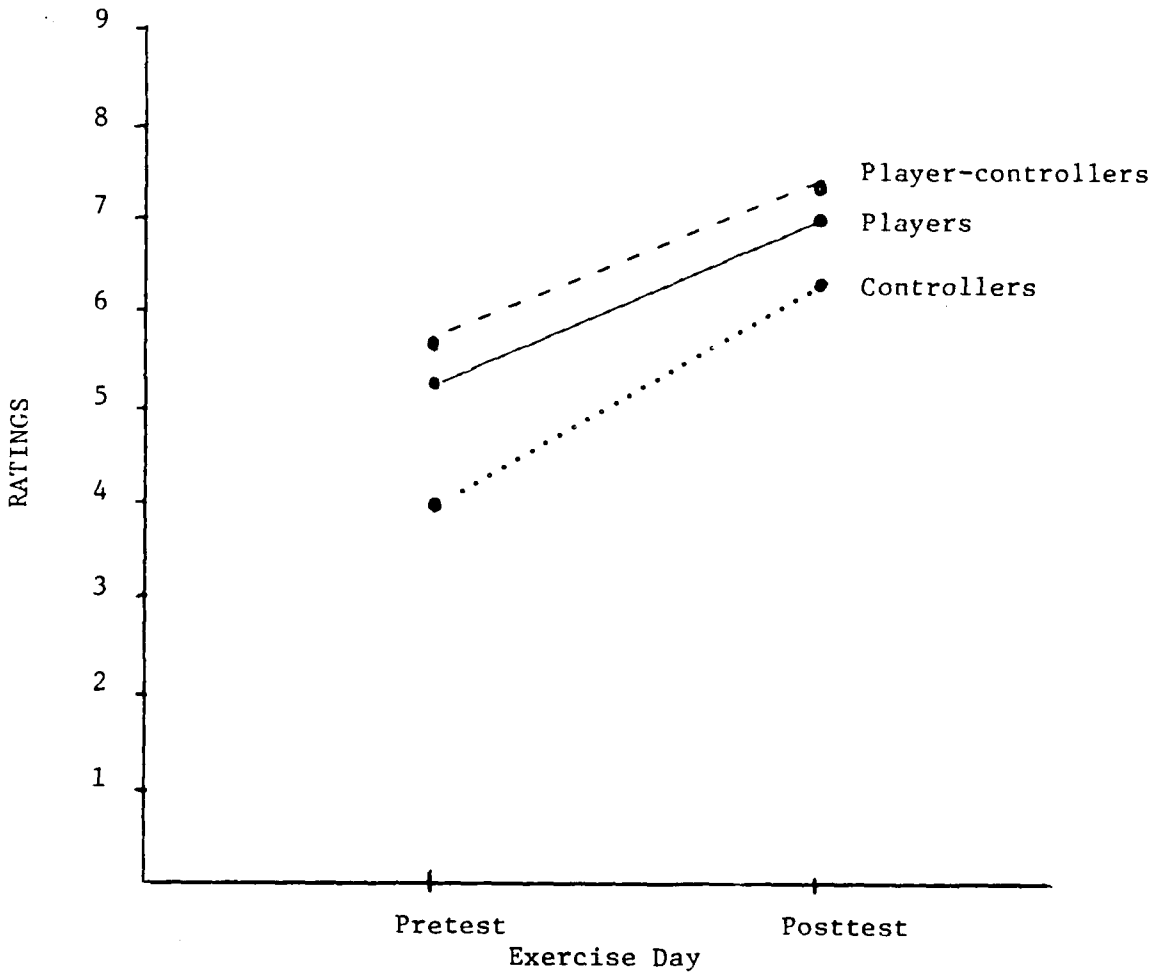



Figure 2. Mean ARTEP performance ratings by controllers, players, and player-controllers on pre- and posttest exercises.

Table 7

Mean Controller Ratings of Units for Both Pre- and Posttest Exercises

| | PRE | POST |
|---|---|---|
| Unit 1 | 3.80 | 5.54 |
| Unit 2 | 3.80 | 6.77 |
| Unit 3 | 4.16 | 6.21 |
| Unit 4 | 4.29 | 7.23 |
| Unit 5 | 3.94 | 6.20 |
| $\bar{X}$ | 4.00 | 6.39 |

planning phase as measured by the information-flow questionnaire, and the execution phase as measured by probes. Battlefield mission accomplishment scores also improved as did ratings of command group performance by the players, controllers, and player-controllers, and ratings of individual member performance by the controllers.

In contrast, a previous experiment (Thomas, Barber, and Kaplan, 1983), in which feedback was limited to a brief after-action review conducted by players and player-controllers at the end of each day, showed little evidence of improvement over a four-day exercise. There was no significant improvement in information-flow scores or in battle simulation outcomes. The players' self-ratings and the controller ratings of command group performance did increase, but the player-controllers' ratings did not. Probes were not used. The major difference between the two experiments was the addition of the diagnostic and feedback modules in the present experiment.

In conclusion, BCG's improved on all measures of performance when systematic diagnostic feedback was provided in addition to CATTS training. For these results to be generalized to ARTBASS, training exercises must be conducted in a manner similar to the procedures used in the current research. Therefore, procedures for performance measurement should be implemented, and feedback sessions conducted by well trained individuals should be included as part of the typical ARTBASS exercise.

## REFERENCES


Army Training and Evaluation Program for Mechanized Infantry/Tank Task Forces, No. 71-2, Washington, D.C.: Headquarters, Department of the Army, 17 June 1977.

Barber H. F. and Kaplan, I. T. Battalion Command Group Performance in Simulated Combat. ARI Technical Paper 353, March 1979. (AD A070 089)

Barber, H. F. and Solick, R. E. MILES Training and Evaluation Test, USAEUR: Battalion Command Group Training. ARI Research Report 1290, June 1980. (AD A109 192)

Kaplan, I. T. and Barber, H. F. Training Battalion Command Groups in Simulated Combat: Identification and Measurement of Critical Performance. ARI Technical Paper 376, June 1979. (AD A075 414)

Thomas, G. S., Barber, H. F., and Kaplan, I. T. The Impact of CATTS System Characteristics on Selected Measures of Battalion Command Group Performance. ARI Technical Report 609, 1983. (AD A140 231)

Thomas, G. S. and Cocklin, T. G. A Model of Mission Accomplishment in Simulated Battle. ARI Technical Report 599, 1983. (AD A148 234)

# APPENDIX A

## Example Questions from the

## Information-Flow Questionnaire

Information Flow

Position _____ Date _____

Answer all questions. Circle the letter before the answer that you believe is correct. Do not guess. If you do not know the answer, answer "unknown." Some questions may cover information to which you did not have access.

Indicate how important it is for you to know each item by assigning it a number from 1 to 9, according to the following scale:

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| useless | | | | moderately important | | | | essential |

1. The ADA weapons control status is _____.

   a. free
   b. tight — Importance _____
   c. hold
   d. unknown

2. Attack helicopters available to your TF for planning purposes is _____.

   a. one platoon
   b. two platoons — Importance _____
   c. one company
   d. unknown

3. Your TF has been instructed to go to BP _____ after passage of lines.

   a. 12
   b. 14 — Importance _____
   c. 8
   d. unknown

4. FM radio listening silence will be imposed until _____.

   a. lifted by TF headquarters
   b. reaching the SP — Importance _____
   c. reaching the RP
   d. unknown

5. The minefields and obstacles emplaced in your sector by the 201st ACR _____ by the engineers.

   a. are planned
   b. are in the process of being verified — Importance _____
   c. have been verified
   d. unknown

# APPENDIX B

## Example Of A Probe

Controller S3

PROBE/OBJECTIVE MEASURE

| Event | Insertion Time | Expected Response | Actual Response |
|---|---|---|---|
| Request SITREPS | Desired ½ hr before end of battle or passage of lines | S3 should examine records and provide Bde with situation report | |
| | Actual | From Whom: Bn S3 | |
| | | To Whom: Bde S3 | |
| | | Timeliness: Immediate | |
| | | Completeness: Front line trace friendly losses probability of completing mission others | |
| | | Accuracy: Compare with computer results. 15 min prior | |

Comments:

# APPENDIX C

## Controller ARTEP Rating Form

Performance Estimate

Position _____ Date _____

Indicate how well you think the command group can perform tasks 1 to 5 below. Also rate the two overall performance items. Assign each task/item a number from 1 to 9, according to the following scale:

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| Poor | | Fair | | Good | | Very Good | | Outstanding |

_____ 1. Gather and analyze required information. Includes: (a) analyze mission, (b) determine what information is available and what additional information is required, (c) determine what information sources are available, (d) gather all available information and request additional information as needed.

_____ 2. Develop a plan based on mission and modify it as required by events. Includes: (a) determine friendly capabilities and limitations, request additional assets if needed, (b) estimate enemy capabilities and likely courses of action, (c) identify key terrain, (d) select battle position/routes to objectives (e) identify critical place, (f) develop and compare courses of action, (g) individual staff planning for communications, intelligence, operations, admin/log, fires, and (h) coordinate with other staff members.

_____ 3. Communicate/coordinate. Includes: (a) issue a warning order, (b) disseminate plans and orders, and (c) disseminate combat information and intelligence to higher and lower.

_____ 4. Implement plan. Includes: (a) concentrate/shift combat power, (b) reinforce terrain.

_____ 5. Supervise combat operations. Includes: (a) compare battlefield events with current order and concept of operations, (b) determine that a new course of action is necessary, and (c) determine that a change in implementation is necessary.

_____ Overall, how well did the command group perform its tasks?

Overall, how well did the staff members perform their own particular tasks?

_____BC

_____S1

_____S2

_____S3

_____S4

_____FSO

_____ALO